

On Reconstructability of Quadratic Utility Functions from the Iterations in Gradient Methods[☆]

Farhad Farokhi^{a,*}, Iman Shames^a, Michael G. Rabbat^b, Mikael Johansson^c

^a*Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Australia*

^b*Department of Electrical and Computer Engineering, McGill University, Montréal, Québec, Canada*

^c*ACCESS Linnaeus Center, Electrical Engineering, KTH Royal Institute of Technology, Stockholm, Sweden*

Abstract

In this paper, we consider a scenario where an *eavesdropper* can read the content of messages transmitted over a network. The nodes in the network are running a gradient algorithm to optimize a quadratic utility function where such a utility optimization is a part of a decision making process by an *administrator*. We are interested in understanding the conditions under which the eavesdropper can reconstruct the utility function or a scaled version of it and, as a result, gain insight into the decision-making process. We establish that if the parameter of the gradient algorithm, i.e., the step size, is chosen appropriately, the task of reconstruction becomes practically impossible for a class of Bayesian filters with uniform priors. We establish what step-size rules should be employed to ensure this.

Keywords: Bayesian Inference, Privacy, Data Confidentiality, Gradient Algorithm

1. Introduction

In recent decades, tremendous advances in the areas of communication and computation have facilitated the construction of complex systems. The design and analysis of these systems involve solving large optimization problems. Utility maximization, optimal flow, expenditure minimization, and traffic optimization are examples of such problems. Due to the size of these problems, it is often required that problems are solved over a network of interconnected processors. In many scenarios, the implementation of the solution to the optimization problem is in the public domain. However, from an operational point of view, it is important

[☆]The work of I. Shames is supported by McKenzie Fellowship. The work of F. Farokhi is supported by the Australian Research Council (LP130100605).

*Corresponding author

Email addresses: `farhad.farokhi@unimelb.edu.au` (Farhad Farokhi), `iman.shames@unimelb.edu.au` (Iman Shames), `michael.rabbat@mcgill.ca` (Michael G. Rabbat), `mikaelj@ee.kth.se` (Mikael Johansson)

that the way that the decision is made remains confidential. In other words, while the optimal decision can be known by everyone, the utility function itself should remain confidential. Portfolios in portfolio optimization and local utilities in resource allocation can be considered as examples of such utility functions that need to be kept confidential. This especially becomes an issue as more computations related to operating the critical infrastructure (e.g. power distribution networks) are carried out in the cloud [1]. The importance of confidentiality, integrity, and availability are well understood in the security of data and ICT services [2] and cloud computing [3]. In these settings, confidentiality corresponds to ensuring the non-disclosure of data, integrity is related to the trustworthiness of data, and availability is concerned with the timely access to the data or system functionalities.

In this paper, we mainly focus on the question of confidentiality—particularly, the confidentiality of the utility functions even when the security of the network is compromised and an eavesdropper can listen to all the information being exchanged over the network during the course of solving the optimization problem. We consider scenarios where the utility function has a quadratic form. Specifically, the following question is answered: *when is it possible to reconstruct a utility function, or a scaled version of it, via having access to the iterations produced by an iterative method?* The iterative method considered in this paper is a gradient ascent algorithm. The choice of a gradient algorithm is inline with the recent observations that cast a favourable light on employing first-order methods to solve very large optimization problems [4]. Note that the choice of quadratic programs is not very restrictive as trust-region optimization techniques allow us to solve any general optimization problem using a sequence of constrained quadratic programs recursively [5].

The problem that is addressed here is related to the one considered in the context of differential privacy [6] and, to a larger extent, the application of differential privacy in optimization [7, 8, 9]. However, it is important to note that there, the price for guaranteed confidentiality is paid in terms of data integrity and the accuracy of the solution. To ensure differential privacy, it is known that the information passed between the processing nodes at each step of the optimization algorithm should be perturbed by a random variable from a Laplace distribution [6]. This results in the algorithm not yielding an accurate solution. Here, we argue that the confidentiality of the objective (but not the solution) can be guaranteed in practice with no impact on the accuracy of the solution, if the algorithm parameter (the step size) is chosen appropriately, i.e., it is picked randomly from a sufficiently large set of suitable step sizes. In addition to differential privacy, other notions of privacy in optimization and machine learning have recently been pursued, e.g., see [10, 11, 12]. Note that, in this paper, we are not directly contributing to the privacy-preserving literature, *per se*. Our main objective is point out that, in the setups discussed, one does not need to worry about

privacy since estimating the underlying parameters is practically impossible due computational restrictions (at least, with current technologies). Note that this problem is also related to the system identification and the parameter estimation literature, where the aim is to extract the parameters of the underlying utility or dynamics. However, in our setup, the eavesdropper cannot inject proper reference signals to fully probe the system (that is commonly known as the *persistent excitation* and is necessary for achieving the estimation objective [13]). Finally, in [14], the agents use their actions to learn about the strategies or the utilities of the other agents to subsequently devise optimal strategies. However, in that study, the computational aspects of the problem were largely unexplored and only linear programs were considered.

The outline of this paper is as follows. In the next section, the problem that is considered in this paper is formulated. In Section 3, we consider the case where the eavesdropper has access to the iterates that are generated during the course of solving an unconstrained quadratic program. In this section, different choices of the step size are considered and conditions for which the utility function cannot be constructed successfully are discussed. Next, in Section 4, we consider the case where the problem is constrained. Concluding remarks are given in Section 5.

1.1. Notation

The sets of reals, nonnegative reals, integers, and nonnegative integers are, respectively, denoted by \mathbb{R} , $\mathbb{R}_{\geq 0}$, \mathbb{Z} , and $\mathbb{Z}_{\geq 0}$. The rest of the sets are denoted by calligraphic Roman letters, such as \mathcal{M} . Specifically, \mathcal{S}_+^n is defined to be the set of symmetric positive-definite matrices in $\mathbb{R}^{n \times n}$. We define $\text{vec} : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{nm}$ to be a vectorization operator that puts all the columns of a matrix into a vector sequentially. Finally, we use $A \otimes B$ to be the Kronecker product of matrices A and B .

2. Problem Formulation

Consider the following optimization problem:

$$\max_{x \in \mathbb{R}^n} \quad -\frac{1}{2}x^\top Qx - q^\top x, \quad (1a)$$

$$\text{s.t.} \quad Cx \leq d, \quad (1b)$$

where $Q \in \mathcal{S}_+^n$, $q \in \mathbb{R}^n$, $C \in \mathbb{R}^{m \times n}$, $d \in \mathbb{R}^m$, and $\mathcal{X} \triangleq \{x \in \mathbb{R}^n \mid Cx \leq d\} \neq \emptyset$. The optimization problem (1) is solved by an *administrator* over a network via an optimization method, $\mathcal{F}(\cdot)$, given by

$$x[k+1] = \mathcal{F}(x[k]), x[0] \in \mathcal{X}. \quad (2)$$

Throughout this paper, we assume that $\mathcal{F}(\cdot)$ is the gradient ascent algorithm in which different step-size selection methods can be used. This assumption is, partly, motivated by favourable results on first-order methods for solving large-scale optimization problems [4]. However, this assumption is also in place to greatly simplify the proofs and the presentation.

Remark 1. *At first glance, the update rule in (2) appears to be a centralized implementation. However, distributed algorithms using primal decomposition as well as the inner problems for distributed algorithms using dual decomposition (see [15]) can be both rewritten, albeit in an aggregated form, in the form of (2).*

Remark 2. *The results presented in this work, at least in part, are applicable to more general utility functions, e.g., logarithmic functions. However, the selection of the quadratic utility functions results in linear operators that greatly simplifies the proofs. Moreover, the quadratic utility functions, although partially conservative, have many applications and are widely used in signal processing, e.g., weighted least squares, and machine learning, e.g. support vector machines (SVM) [11].*

The measurement model of the eavesdropper is as follows. For any two consecutive measurements of the optimization variable $x[k]$ and $x[k+1]$, for some $k \in \mathbb{Z}_{\geq 0}$, the eavesdropper can construct a measurement of the form

$$y[k] = x[k] - x[k+1]. \quad (3)$$

Therefore, at time step $k+1$, the eavesdropper has access to measurement pairs $(x[t], y[t])_{t=0}^k$. Providing the solution to the following problems is of interest.

Problem 1 (Utility Function Reconstruction). *Assuming that the eavesdropper can measure $x[k]$ for all k and the values of A and b are known, under what conditions on the step size selection of the gradient descent algorithms can the eavesdropper estimate (\hat{Q}, \hat{q}) such that $Q = \gamma \hat{Q}$ and $q = \gamma \hat{q}$ for some $\gamma > 0$?*

Solving the problem above enables the eavesdropper to determine the way that decisions are made. For example, it can be determined which variable has a bigger impact on the solution of the optimization problem (1). Hence, it is not necessary to exactly estimate γ .

Remark 3. *In this paper, we assume that the communication is carried out over real and noiseless channels. Alternatively, one may consider the effects of quantisation and noise on the utility reconstruction problem. However, this is beyond the scope of this paper.*

Finally, we have the following standing assumption.

Assumption 1. *The parameters $(Q, q) \in \mathcal{Q} \subseteq \mathcal{S}_+^n \times \mathbb{R}^n$ are randomly generated according to the non-degenerate probability density function $p : \mathcal{Q} \rightarrow \mathbb{R}_{\geq 0}$. Further, we assume the distribution of (Q, q) is independent of the initialization of the algorithm $x[0]$, which is uniformly selected from $\{x | x^\top x \leq 1\}$. The eavesdropper knows these probability distributions.*

In the following sections, first we consider the case where the problem (1) is unconstrained, and then we study the constrained case. Note that the choice of unconstrained quadratic programs is not restrictive. We can solve any general optimization problem using a sequence of constrained quadratic programs recursively using trust-region optimization techniques. Further, when using primal-dual techniques, we can solve a constrained quadratic program recursively using a sequence of unconstrained quadratic programs. Alternatively, as also discussed in Section 4, we can use logarithmic barrier functions to solve a constrained quadratic program.

3. Unconstrained Case

In the case where the optimization problem is unconstrained, the gradient iterations are such that

$$x[k+1] = x[k] - \alpha[k](Qx[k] + q), \quad (4)$$

where $\alpha[k]$ is an appropriately selected step size (e.g., it is well known that if $\alpha[k], \forall k \in \mathbb{Z}_{\geq 0}$, belongs to an appropriately selected interval on the positive reals, the iterations in (4) converge to the optimal solution [16]).

Remark 4. *As remarked earlier, in this paper our primary interest is in the case where the gradient iterations are implemented in a distributed manner. For instance, if we employ n processors, each processor needs to follow the update rule*

$$x_i[k+1] = (1 - \alpha[k]q_{ii})x_i[k] - \sum_{j \neq i} \alpha[k]q_{ij}x_j[k] - \alpha[k]q_i,$$

where $x_i[k]$ is i 'th element of the decision vector $x[k]$. To implement this update rule, the processors need to communicate the elements of the decision vector over the directed graph \mathcal{G} with vertex set $\mathcal{V}_{\mathcal{G}} = \{1, \dots, n\}$ and edge set $\mathcal{E}_{\mathcal{G}} = \{(i, j) | 1 \leq i \neq j \leq n, q_{ij} \neq 0\}$. The messages that the processors pass contain $x_i[k], \forall i$, and by observing these messages, the eavesdropper can obtain $(x[t])_{t \in \mathbb{Z}_{\geq 0}}$. As a viable avenue for future research, we can consider the scenario in which the eavesdropper can only listen to a subset of the transmitted messages.

In this scenario, the eavesdropper measurement model given by (3) becomes

$$y[k] = \alpha[k](Qx[k] + q).$$

As before, at time step $k + 1$, measurement pairs $(x[t], y[t])_{t=0}^k$ are available to the eavesdropper. We make the following standing assumption.

Assumption 2. *The vectors $(x[t])_{t=0}^{n-1}$ are linearly independent.*

We formally characterize the conditions for which Assumption 2 holds in the course of this paper for each scenario.

Remark 5. *Under Assumption 1, the independence assumption is without loss of generality (i.e., Assumption 2 holds almost surely). Note that when using the gradient ascent algorithm, for the measurements to be dependent, the algorithm should be initialized at a point along one of the eigenvectors of Q , and q should be parallel to that specific eigenvector.*

Remark 6. *Let us briefly explain why Assumption 2 is necessary. In this remark, we assume that $\alpha[k]$, $k \in \mathbb{N}$, is known. Note that this shows the necessity of Assumption 2 as it discusses a more restrictive setup (because the eavesdropper has access to more information). In such case, we have*

$$\begin{aligned} Qx[k] + q &= \text{vec}(Qx[k] + q) \\ &= \text{vec}(Qx[k]) + q \\ &= (x[k]^\top \otimes I) \text{vec}(Q) + q, \end{aligned}$$

where the last equality follows from Item (5) in [17, p. 97]. This gives

$$\left(\begin{bmatrix} x[0]^\top & 1 \\ \vdots & \vdots \\ x[k]^\top & 1 \end{bmatrix} \otimes I \right) \begin{bmatrix} \text{vec}(Q) \\ q \end{bmatrix} = \begin{bmatrix} y[0]/\alpha[0] \\ \vdots \\ y[k]/\alpha[k] \end{bmatrix}. \quad (5)$$

Let us denote the matrix on the left hand side of (5) by G . To avoid admitting redundant equations, G should have a full row rank. From the properties of the Kronecker product [17, p. 58], if Assumption 2 holds, for all $k \leq n - 1$, we know that

$$\text{rank} \left(\begin{bmatrix} x[0]^\top & 1 \\ \vdots & \vdots \\ x[k]^\top & 1 \end{bmatrix} \otimes I \right) = \text{rank} \left(\begin{bmatrix} x[0]^\top & 1 \\ \vdots & \vdots \\ x[k]^\top & 1 \end{bmatrix} \right) \text{rank}(I) = (k + 1)n.$$

The number of rows of G is also equal to $(k + 1)n$ and, hence, G has full row rank. Consequently, there is no redundant equation.

3.1. Constant Step Size

In this subsection, we assume $\alpha[k] = \alpha > 0$ for all $k \in \mathbb{Z}_{\geq 0}$. In this case, it is not possible to reconstruct Q, q uniquely because α shows itself as a scaling factor in these matrices. In other words, for γ in Problem 1 we have $\gamma = 1/\alpha$. Let us construct the set $\mathcal{M}[k] = \{(Q', q') \in \mathcal{S}_+^n \times \mathbb{R}^n \mid y[t] = Q'x[t] + q', \forall t = 0, \dots, k\}$. This denotes the set of parameters that are consistent with our measurements up to time step $k + 1$ (note that for constructing $(y[t])_{t=0}^k$, the eavesdropper needs to measure $(x[t])_{t=0}^{k+1}$). Evidently, by construction, these sets are nonexpansive, i.e., $\mathcal{M}[k + 1] \subseteq \mathcal{M}[k]$ for any $k \in \mathbb{Z}_{\geq 0}$. First, let us present a condition under which Assumption 2 holds.

Remark 7. *The introduced estimator $\mathcal{M}[k]$ has close connections to the idea behind set-membership identification in which the set of permissible parameters are gradually reduced by removing realisations that are not compatible with the newly received measurements; see [18, 19, 20].*

Lemma 1. *Let the distribution p governing Q (cf. Assumption 1) be such that the algebraic multiplicity of every eigenvalue of Q is almost surely equal to one. Then, $(x[t])_{t=0}^{n-1}$ are almost surely independent.*

Proof. Notice that

$$x[k + 1] = x[k] - \alpha(Qx[k] + q) = (I - \alpha Q)x[k] - \alpha q.$$

Therefore, we have

$$\begin{bmatrix} x[0]^\top \\ x[1]^\top \\ \vdots \\ x[n-1]^\top \end{bmatrix} = \begin{bmatrix} x[0]^\top \\ x[0]^\top (I - \alpha Q)^\top \\ \vdots \\ x[0]^\top (I - \alpha Q)^{(n-1)\top} \end{bmatrix} - \begin{bmatrix} 0_{1 \times n} \\ (\alpha q)^\top \\ \vdots \\ \sum_{j=0}^{n-2} (\alpha q)^\top (I - \alpha Q)^j \end{bmatrix}. \quad (6)$$

Since $x[0]$ is selected randomly and independently from the pair (Q, q) , the iterates $(x[t])_{t=0}^{n-1}$ are independent (i.e., the matrix on the left-hand side of (6) is full rank) if

$$\text{rank} \left(\begin{bmatrix} x[0]^\top \\ x[0]^\top (I - \alpha Q)^\top \\ \vdots \\ x[0]^\top (I - \alpha Q)^{(n-1)\top} \end{bmatrix} \right) = n, \quad (7)$$

which is equivalent to that the pair $((I - \alpha Q)^\top, x[0]^\top)$ is observable. From the controllability/observability literature [21, p. 123], we know that (7) holds if and only if

$$\text{rank} \left(\begin{bmatrix} (I - \alpha Q)^\top - \mu I \\ x[0]^\top \end{bmatrix} \right) = n$$

for all eigenvalues μ of $(I - \alpha Q)^\top$. This condition is satisfied if (i) the algebraic multiplicity of all the eigenvalues of $(I - \alpha Q)^\top$ is equal to one [17, p. 59] and (ii) $x[0]$ is not an eigenvector of $(I - \alpha Q)^\top$ (which is satisfied with probability one since $x[0]$ and Q are drawn independently). This is, in turn, satisfied if the algebraic multiplicity of every eigenvalue of Q is equal to one. \square

The following theorem shows that after enough measurements, the set $\mathcal{M}[k]$ becomes a singleton.

Theorem 1. *Let $n \geq 3$. Then $\mathcal{M}[k] = \{(\alpha Q, \alpha q)\}$ for all $k \geq \lceil (n+1)/2 \rceil$.*

Proof. Because Q' is a symmetric matrix, it only has $(n^2 + n)/2$ unknowns ($= 1 + 2 + \dots + n$). Therefore, the eavesdropper needs to calculate $(n^2 + 3n)/2$ unknowns corresponding with the entries of Q' and q' . Due to Assumption 2, if the eavesdropper collects measurements for up to $k = \lceil (n+1)/2 \rceil$ (notice that, by definition, the number of measurements in $\mathcal{M}[k]$ is equal to $k + 1$), the set of linear equations defining $\mathcal{M}[k]$ admits a unique solution, (\hat{Q}, \hat{q}) , where $\hat{Q} = \alpha Q$ and $\hat{q} = \alpha q$. To be able to use Assumption 2, we should have $\lceil (n+1)/2 \rceil = k \leq n - 1$ based on Remark 6, which gives $n \geq 3$. \square

Remark 8. *If the eavesdropper collects $k < \lceil (n+1)/2 \rceil$ measurements, the set $\mathcal{M}[k]$ has infinitely many elements. As a result, if the iterations are terminated at $k < \lceil (n+1)/2 \rceil$ iterations, it is impossible for the eavesdropper to reconstruct the parameters of the utility function. However, the confidentiality is guaranteed here at the price of getting a possibly inaccurate solution.*

Remark 9. *For $n < 3$, regardless of the number of collected measurements, $\mathcal{M}[k]$ never becomes a singleton. This is due to the fact that Assumption 2 does not hold any more.*

Remark 10. *The presented estimator also works if the step sizes are selected as $\alpha[k] = c/k^\delta$ for all $k \in \mathbb{Z}_{\geq 0}$ and for a fixed $\delta \in (1/2, 1]$. This is true because we can always scale the measurements $(x[t], y[t])_{t=0}^k$ to $(x[t], t^\delta y[t])_{t=0}^k$ and, subsequently, use the presented results for the constant step size. If δ is not known by the eavesdropper, we can construct a filter to also reconstruct δ based on the measurements $(x[t], y[t])_{t=0}^k$, however, this would make the problem considerably more difficult because of the complexity of the Bayes updates in this case.*

This remark shows that the choice of a time-varying step size as described above does not make the reconstruction of the utility function any harder than the constant step size case so long as δ is publicly known.

3.2. Random Step Sizes Drawn from a Finite Set

Next we consider the case where the step sizes $\alpha[k]$, for $k \in \mathbb{Z}_{\geq 0}$, are drawn uniformly from a set $\mathcal{A} = \{\alpha^{(1)}, \dots, \alpha^{(s)}\}$ of s distinct values. Moreover, we assume that the step sizes $(\alpha[k])_{k \in \mathbb{Z}_{\geq 0}}$ are independently and identically distributed over time. First, let us present a condition for which Assumption 2 holds.

Lemma 2. *Let \mathcal{Q} be such that the following condition is almost surely satisfied*

$$\text{rank} \left(\begin{bmatrix} x[0]^\top \\ x[0]^\top (I - \alpha[0]Q)^\top \\ \vdots \\ x[0]^\top (I - \alpha[0]Q)^\top \cdots (I - \alpha[n-2]Q)^\top \end{bmatrix} \right) = n. \quad (8)$$

Then, $(x[t])_{t=0}^{n-1}$ are almost surely independent.

Proof. Similar to the proof of Lemma 1, notice that $x[k+1] = (I - \alpha[k]Q)x[k] - \alpha[k]q$. Therefore, we have

$$\begin{aligned} \begin{bmatrix} x[0]^\top \\ x[1]^\top \\ \vdots \\ x[n-1]^\top \end{bmatrix} &= \begin{bmatrix} x[0]^\top \\ x[0]^\top (I - \alpha[0]Q)^\top \\ \vdots \\ x[0]^\top (I - \alpha[0]Q)^\top \cdots (I - \alpha[n-2]Q)^\top \end{bmatrix} \\ &\quad - \begin{bmatrix} 0_{1 \times n} \\ (\alpha q)^\top \\ \vdots \\ \sum_{j=0}^{n-2} (\alpha[j]q)^\top (I - \alpha[j+1]Q)^\top \cdots (I - \alpha[n-2]Q)^\top \end{bmatrix}. \end{aligned} \quad (9)$$

Since $x[0]$ is selected randomly and independently from the pair (Q, q) , the iterates $(x[t])_{t=0}^{n-1}$ are independent if the condition in the statement of the lemma holds. \square

Remark 11. *Condition (8) is intimately related to the observability of time-varying linear systems $((I - \alpha[k]Q)^\top, x[0]^\top)$; see [22, p. 462].*

At iteration $k + 1$, after observing $(x[t], y[t])_{t=0}^k$, the eavesdropper may use the Bayes' update rule, consecutively, for generating the conditional density function

$$\begin{aligned} p(Q', q' | (x[t], y[t])_{t=0}^k) &\propto p(y[k] | Q', q', (x[t], y[t])_{t=0}^{k-1}, x[k]) p(Q', q' | (x[t], y[t])_{t=0}^{k-1}, x[k]) \\ &= p(y[k] | Q', q', (x[t], y[t])_{t=0}^{k-1}, x[k]) p(Q', q' | (x[t], y[t])_{t=0}^{k-1}) \end{aligned} \quad (10a)$$

$$= p(y[k] | Q', q', x[k]) p(Q', q' | (x[t], y[t])_{t=0}^{k-1}) \quad (10b)$$

$$= \left[\sum_{\alpha' \in \mathcal{A}} p(y[k] | Q', q', x[k], \alpha') p(\alpha') \right] p(Q', q' | (x[t], y[t])_{t=0}^{k-1}) \quad (10c)$$

$$\propto \left[\sum_{\alpha' \in \mathcal{A}} p(y[k] | Q', q', x[k], \alpha') \right] p(Q', q' | (x[t], y[t])_{t=0}^{k-1}) \quad (10d)$$

where (10a) follows from independence of the cost parameters (Q, q) and $x[k]$ given $(x[t], y[t])_{t=0}^{k-1}$ (note that $x[k]$ is merely the negation of the summation of all $(y[t])_{t=0}^{k-1}$ plus $x[0]$ and, hence, it is redundant information), (10b) follows from independence of $y[k]$ from $(x[t], y[t])_{t=0}^{k-1}$ given $x[k]$ and the cost parameters (Q, q) , (10c) follows from conditioning on α , and (10d) follows from the uniform distribution of the step sizes.

Now, note that

$$\begin{aligned} \sum_{\alpha' \in \mathcal{A}} p(y[k] | Q', q', x[k], \alpha') &= \begin{cases} 1, & \exists \alpha'' \in \mathcal{A} : y[k] = \alpha''(Q'x[k] + q'), \\ 0, & \text{otherwise,} \end{cases} \\ &= \mathbb{1}_{(Q', q') \in \mathcal{D}(x[k], y[k])}, \end{aligned}$$

where $\mathcal{D}(x[k], y[k]) = \{(Q', q') \in \mathcal{S}_+^n \times \mathbb{R}^n \mid \exists \alpha' \in \mathcal{A} : y[k] = \alpha'(Q'x[k] + q')\}$. Hence, we have $p(Q', q' | (x[t], y[t])_{t=0}^k) \propto \mathbb{1}_{(Q', q') \in \mathcal{D}(x[k], y[k])} p(Q', q' | (x[t], y[t])_{t=0}^{k-1})$. Using induction, we can show that

$$p(Q', q' | (x[t], y[t])_{t=0}^k) \propto \left[\prod_{t=0}^k \mathbb{1}_{(Q', q') \in \mathcal{D}(x[t], y[t])} \right] p(Q', q') = \mathbb{1}_{(Q', q') \in \cap_{t=0}^k \mathcal{D}(x[t], y[t])} p(Q', q').$$

Now, we can redefine

$$\mathcal{M}[k] = \cap_{t=0}^k \mathcal{D}(x[t], y[t]) = \{(Q', q') \in \mathbb{R}^{n \times n} \times \mathbb{R}^n \mid \exists \alpha'[t] \in \mathcal{A} : y[t] = \alpha'[t](Q'x[t] + q'), \forall t = 0, \dots, k\}.$$

Therefore, Bayes' rule gives $p(Q', q' | (x[t], y[t])_{t=0}^k) = \mathbb{1}_{(Q', q') \in \mathcal{M}[k]} p(Q', q')$. The next theorem shows that the set $\mathcal{M}[k]$ becomes a singleton and, hence, the Bayesian filter converges to the correct parameter selection.

Theorem 2. *Let $n \geq 5$. Then $\mathcal{M}[k] = \{(Q, q)\}$ for all $k \geq \lceil (n + 3)/2 \rceil$.*

Proof. Let us enumerate all the possible sequences of the step sizes \mathcal{A}^{k+1} for each k . Now, for any sequence of step sizes $(\alpha'[t])_{t=0}^k \in \mathcal{A}^{k+1}$, the consistent parameter sets are given by the set-valued mapping

$\mathcal{Z}[k; (\alpha'[t])_{t=0}^k] = \{(Q'x[t] + q') = y[t]/\alpha'[t], \forall t = 0, \dots, k\}$. Evidently, $\mathcal{Z}[k; (\alpha'[t])_{t=0}^k]$ is either a singleton or the empty set if $(n+1)/2 \leq k$. This is the case because for the mentioned horizon length, given the sequences of the step sizes, the number of the equations (which we assumed are not redundant; see Assumption 2) is larger than or equal the number of free variables. Now, if we get only one more measurement, i.e., $k = \lceil (n+3)/2 \rceil$, only one of these sets remain nonempty and that points to the true parameters. To be able to use Assumption 2, we should have $\lceil (n+3)/2 \rceil = k \leq n-1$, which gives $n \geq 5$. \square

Note that the estimator constructed in the proof of Theorem 2 relies on the fact that we can enumerate all the possible sequences of the step sizes for $k = \lceil (n+3)/2 \rceil$ and solve a set of linear equations for each one to extract the true parameters. The number of all the possible sequences of the step sizes is equal to $s^{\lceil (n+3)/2 \rceil} = \mathcal{O}(s^n)$. Thus, this estimator is practically implementable only for relatively small s and n . However, this problem can be fixed with a simple change of variable. To do so, we can alternatively define the set of utility functions consistent with the observations as

$$\mathcal{M}[k] = \{(Q', q') \in \mathbb{R}^{n \times n} \times \mathbb{R}^n \mid \exists \beta[t] \in \{1/\alpha^{(1)}, \dots, 1/\alpha^{(s)}\} : \beta[t]y[t] = Q'x[t] + q', \forall t = 0, \dots, k\}.$$

Following the same line of reasoning as in Remark 6, we can see that the elements of $\mathcal{M}[k]$ are the solutions of the set of equations

$$\left(\begin{bmatrix} x[0]^\top & 1 \\ \vdots & \vdots \\ x[k]^\top & 1 \end{bmatrix} \otimes I \right) \begin{bmatrix} \text{vec}(Q') \\ q' \end{bmatrix} = \begin{bmatrix} y[0]\beta[0] \\ \vdots \\ y[k]\beta[k] \end{bmatrix}.$$

We may rewrite this set of equations as

$$\underbrace{\begin{bmatrix} x[0]^\top \otimes I & I & y[0] & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x[k]^\top \otimes I & I & 0 & \cdots & y[k] \end{bmatrix}}_{:= \Phi} \begin{bmatrix} \text{vec}(Q') \\ q' \\ \beta[0] \\ \vdots \\ \beta[k] \end{bmatrix} = 0.$$

Using the arguments of Remark 6 and Lemma 1, we may observe that Φ is a full row rank matrix. Now, note that the number of unknown decision variables here is $k + n(n+1)/2 + n$ and the number of equations is $(k+1)n$. Therefore, for this set of equations to have a unique solution (up to a scaling), we need to have $k + n(n+1)/2 + n \leq (k+1)n$ which means $k \geq n(n+1)/(2(n-1))$. This is satisfied if we select

$k \geq \lceil (n+3)/2 \rceil$ as recommended in Theorem 2. Therefore, with this change of variable, we can reconstruct the utility function based on the observations in polynomial time.

Remark 12. *Note that, in this subsection, we did not use the fact that \mathcal{A} is a finite set. Therefore, the presented argument is also valid when the step sizes are selected from closed interval of the positive reals.*

3.3. Agent-Dependent Step Sizes

In this subsection, we assume that each entry of the decision variable $x[k]$ is updated by an agent that can select its step size independently. In this case, the gradient iterations for each entry of the decision variable becomes

$$x_i[k+1] = x_i[k] - \alpha_i[k] \left(q_{ii}x_i[k] + \sum_{j \neq i} q_{ij}x_j[k] + q_i \right),$$

where $\alpha_i[k]$ are independently and identically distributed discrete random variables selected with equal probability from the set \mathcal{A} . Let $\alpha_i[k]$ and $\alpha_j[k]$ be statistically independent if $i \neq j$. As a result, we get

$$x[k+1] = x[k] - A[k](Qx[k] + q),$$

where

$$A[k] = \begin{bmatrix} \alpha_1[k] & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \alpha_n[k] \end{bmatrix}.$$

We may define $\mathcal{A}_d^{n \times n}$ as the set of diagonal matrices of size $n \times n$ with entries belonging to \mathcal{A} . By definition, $A[k] \in \mathcal{A}_d^{n \times n}$ for all $k \geq 0$. Note that we can guarantee the convergence of the gradient algorithm even when using agent-dependent step sizes; see Appendix A for more information. Similar to the previous subsections, let us define

$$y[k] = A[k](Qx[k] + q).$$

As before, at time step $k+1$, measurement pairs $(x[t], y[t])_{t=0}^k$ are available to the eavesdropper. Following the same line of reasoning as in the previous subsections, at iteration $k+1$, the eavesdropper may use the Bayes' update rule, consecutively, for generating the conditional density function

$$p(Q', q' | (x[t], y[t])_{t=0}^k) \propto \mathbb{1}_{(Q', q') \in \mathcal{D}(x[k], y[k])} p(Q', q' | (x[t], y[t])_{t=0}^{k-1}). \quad (11)$$

where $\mathcal{D}(x[k], y[k]) = \{(Q', q') \in \mathcal{S}_+^n \times \mathbb{R}^n \mid \exists A' \in \mathcal{A}_d^{n \times n} : y[k] = A'(Q'x[k] + q')\}$. Hence, using induction, we get

$$p(Q', q' \mid (x[t], y[t])_{t=0}^k) \propto \left[\prod_{t=0}^k \mathbb{1}_{(Q', q') \in \mathcal{D}(x[t], y[t])} \right] p(Q', q') = \mathbb{1}_{(Q', q') \in \mathcal{M}[k]} p(Q', q'),$$

where

$$\mathcal{M}[k] = \cap_{t=0}^k \mathcal{D}(x[t], y[t]) = \{(Q', q') \in \mathbb{R}^{n \times n} \times \mathbb{R}^n \mid \exists A'[t] \in \mathcal{A}_d^{n \times n} : y[t] = A'[t](Q'x[t] + q'), \forall t = 0, \dots, k\}.$$

The next theorem shows Bayesian filter is always inconclusive.

Theorem 3. *The cardinality of the set $\mathcal{M}[k]$ is uncountably infinite for all $k \in \mathbb{Z}_{\geq 0}$.*

Proof. Firstly, note that we can redefine $\mathcal{M}[k]$ as

$$\mathcal{M}[k] = \cap_{t=0}^k \mathcal{D}(x[t], y[t]) = \{(Q', q') \in \mathbb{R}^{n \times n} \times \mathbb{R}^n \mid \exists B[t] \in \mathcal{B}_d^{n \times n} : B[k]y[t] = Q'x[t] + q', \forall t = 0, \dots, k\},$$

where $\mathcal{B}_d^{n \times n}$ denotes the set of all diagonal matrices of size $n \times n$ with diagonal entries belonging to $\mathcal{B} = \{1/\alpha^{(1)}, \dots, 1/\alpha^{(s)}\}$. This way, as described in the previous subsection, we need to solve a linear set of equations to find the entries of the set $\mathcal{M}[k]$. Now, note that, in each iteration, we receive n new measurements while adding n new variables (i.e., step sizes). Therefore, even if all the measurements are independent, there is not a unique solution (since the number of unknowns is always strictly larger than the number of measurements). \square

Theorem 3 shows that no matter how many measurements the eavesdropper gathers, it is impossible to reconstruct the cost function.

4. Constrained Case

For the constrained optimization problem, we add the constraints using logarithmic barrier functions. In that case, we get the unconstrained optimization problem

$$\max_{x \in \mathbb{R}^n} -\frac{1}{2}x^\top Qx - q^\top x + \lambda \sum_{i=1}^m \log(d_i - C_i x),$$

where $C_i \in \mathbb{R}^{1 \times n}$ is i -th row of the matrix C and $\lambda > 0$ is a scaling factor. As λ approaches zero, the solution of this problem converges to the solution of the original constrained optimization problem [23, p. 566]. When

using the gradient algorithm, in this case, we get

$$x[k+1] = x[k] - \alpha[k] \left(Qx[k] + q + \sum_{i=1}^m \frac{\lambda}{d_i - C_i x[k]} C_i^\top \right), \quad x[0] \in \mathcal{X}.$$

Therefore, an eavesdropper that listens to consecutive iterations can construct the measurements

$$y[k] = \alpha[k] \left(Qx[k] + q + \sum_{i=1}^m \frac{\lambda}{d_i - C_i x[k]} C_i^\top \right).$$

In the reminder of this section, we assume that the eavesdropper knows the parameters of the constraints (C, d) and only wants to estimate the parameters of the utility function and the scaling factor.

Remark 13. *In many problems, the constraints are enforced by the physical characteristics of the problem and are hence public knowledge. This is different from the utility function that is typically motivated by the internal mechanism of the system and the priorities of its operator and is hence kept private.*

Let us consider the case that the step sizes are selected randomly and uniformly from a finite set $\mathcal{A} = \{\alpha^{(1)}, \dots, \alpha^{(s)}\}$ as in Subsection 3.2; the proofs for the other cases are not different.

Assumption 3. *The parameter $\lambda \in \Lambda \subset \mathbb{R}_{\geq 0}$ is randomly generated according to the probability density function $p : \Lambda \rightarrow \mathbb{R}_{\geq 0}$. Assume that $x[0]$ is chosen uniformly at random from $\{x | Cx < d\}$. Further, the distribution of λ is independent of the initialization of the algorithm $x[0]$ and utility function parameters (Q, q) .*

Similarly, we can present the following condition for the satisfaction of Assumption 2.

Lemma 3. *Let Q be such that the following condition is almost surely satisfied*

$$\text{rank} \left(\begin{bmatrix} x[0]^\top \\ x[0]^\top (I - \alpha[0]Q)^\top \\ \vdots \\ x[0]^\top (I - \alpha[0]Q)^\top \cdots (I - \alpha[n-2]Q)^\top \end{bmatrix} \right) = n.$$

Then, $(x[t])_{t=0}^{n-1}$ are almost surely independent.

Proof. The proof follows from the same line of reasoning as in Lemma 2. The only difference is that, in this case, the right-hand side of (9) admits additional nonlinear terms that are multiplied by λ . However, since λ is selected independently of the parameters and the initial condition (see Assumption 3), these terms almost surely do not contribute to the rank of the matrix on the left-hand side of (9). \square

At iteration $k + 1$, after observing $(x[t], y[t])_{t=0}^k$, the eavesdropper may use the Bayes' update rule, consecutively, for generating the conditional density function

$$\begin{aligned} p(Q', q', \lambda' | (x[t], y[t])_{t=0}^k) &= \left[\sum_{\alpha' \in \mathcal{A}} p(y[k] | Q', q', \lambda', x[k], \alpha') p(\alpha') \right] p(Q', q', \lambda' | (x[t], y[t])_{t=0}^{k-1}) \\ &\propto \left[\sum_{\alpha' \in \mathcal{A}} p(y[k] | Q', q', \lambda', x[k], \alpha') \right] p(Q', q', \lambda' | (x[t], y[t])_{t=0}^{k-1}). \end{aligned}$$

Similarly, we have $\sum_{\alpha' \in \mathcal{A}} p(y[k] | Q', q', \lambda', x[k], \alpha') = \mathbb{1}_{(Q', q', \lambda') \in \mathcal{D}(x[k], y[k])}$, where

$$\mathcal{D}(x[k], y[k]) = \left\{ (Q', q', \lambda') \in \mathcal{S}_+^n \times \mathbb{R}^n \times \Lambda \mid \exists \alpha' \in \mathcal{A} : y[k] = \alpha' \left(Q' x[k] + q' + \lambda' \sum_{i=1}^m \frac{1}{d_i - C_i x[k]} C_i^\top \right) \right\}.$$

Again, using induction, we can show that

$$p(Q', q', \lambda' | (x[t], y[t])_{t=0}^k) \propto \left[\prod_{t=0}^k \mathbb{1}_{(Q', q', \lambda') \in \mathcal{D}(x[t], y[t])} \right] p(Q', q') p(\lambda') = \mathbb{1}_{(Q', q', \lambda') \in \cap_{t=0}^k \mathcal{D}(x[t], y[t])} p(Q', q') p(\lambda').$$

Now, we can define

$$\begin{aligned} \mathcal{M}[k] &= \cap_{t=0}^k \mathcal{D}(x[t], y[t]) \\ &= \left\{ (Q', q', \lambda') \in \mathcal{S}_+^n \times \mathbb{R}^n \times \Lambda \mid \exists \alpha'[t] \in \mathcal{A} : y[t] = \alpha'[t] \left(Q' x[t] + q' + \lambda' \sum_{i=1}^m \frac{1}{d_i - C_i x[t]} C_i^\top \right), \forall t = 0, \dots, k \right\}, \end{aligned}$$

which gives $p(Q', q', \lambda' | (x[t], y[t])_{t=0}^k) = \mathbb{1}_{(Q', q', \lambda') \in \mathcal{M}[k]} p(Q', q') p(\lambda')$. The next theorem shows that the Bayesian filter converges to the correct parameter selection.

Theorem 4. *Let $n \geq 6$. $\mathcal{M}[k] = \{(Q, q, \lambda)\}$ for all $k \geq \lceil (n+3)/2 + 1/n \rceil$.*

Proof. Let us assume that we use an estimator that, for each time step k , enumerates all the possible sequences of the step sizes \mathcal{A}^{k+1} . Now, for any sequence of step sizes $(\alpha'[t])_{t=0}^k \in \mathcal{A}^{k+1}$, the consistent parameter sets are given by the mapping $\mathcal{Z}[k; (\alpha'[t])_{t=0}^k] = \{(Q' x[t] + q' + \sum_{i=1}^m \frac{\lambda'}{d_i - C_i x[t]} C_i^\top) = y[t] / \alpha'[t], \forall t = 0, \dots, k\}$. Similar to the proof of Theorem 2, $\mathcal{Z}[k; (\alpha'[t])_{t=0}^{k-1}]$ is either a singleton or the empty set if $k+1 \geq ((n+1)n/2 + n+1)/n = (n+3)/2 + 1/n$. This is true since, given the sequences of the step sizes, the number of the equations is larger than or equal the number of free variables. Now, if we get one more measurement, i.e., $k = \lceil (n+3)/2 + 1/n \rceil$, only one of these sets remain nonempty. To be able to use Assumption 2, we should have $\lceil (n+3)/2 + 1/n \rceil = k \leq n-1$, which gives $n \geq 6$. \square

Remark 14. *In the interior point method, the algorithm automatically shrinks the scaling factor λ to extract the optimal point. If the rule for shrinking the scaling factor is known, one can use the Bayesian filter above to estimate the parameters of the utility function.*

Note that, similar to the previous section, we may introduce change of variables to reconstruct the set $\mathcal{M}[k]$, and subsequently the utility function, in polynomial time. To do so, note that we may alternatively define $\mathcal{M}[k]$ as

$$\mathcal{M}[k] = \left\{ (Q', q', \lambda') \in \mathcal{S}_+^n \times \mathbb{R}^n \times \Lambda \mid \exists \beta[t] \in \{1/\alpha^{(1)}, \dots, 1/\alpha^{(s)}\} : \right. \\ \left. \beta[t]y[t] = Q'x[t] + q' + \lambda' \sum_{i=1}^m \frac{1}{d_i - C_i x[t]} C_i^\top, \forall t = 0, \dots, k \right\}.$$

This way, we get a set of linear equations, which as showed in Theorem 4 admits a unique solution for $k \geq \lceil (n+3)/2 + 1/n \rceil$.

Alternatively, we can use independently selected random step sizes at each agent to render the problem of reconstructability impossible. In this case, the update gradient algorithm becomes

$$x[k+1] = x[k] - A[k] \left(Qx[k] + q + \sum_{i=1}^m \frac{\lambda}{d_i - C_i x[k]} C_i^\top \right), x[0] \in \mathcal{X},$$

where $A[k] \in \mathcal{A}_d^{n \times n}$ contains the stochastically-varying agent-dependent step sizes. Therefore, an eavesdropper that listens to consecutive iterations can construct the measurements

$$y[k] = A[k] \left(Qx[k] + q + \sum_{i=1}^m \frac{\lambda}{d_i - C_i x[k]} C_i^\top \right).$$

Now, employing the Bayesian filter, the estimator can deduce that

$$p(Q', q', \lambda' | (x[t], y[t])_{t=0}^k) \propto \mathbb{1}_{(Q', q', \lambda') \in \mathcal{M}[k]} p(Q', q') p(\lambda'),$$

where

$$\mathcal{M}[k] = \left\{ (Q', q', \lambda') \in \mathcal{S}_+^n \times \mathbb{R}^n \times \Lambda \mid \exists A[t] \in \mathcal{A}_d^{n \times n} : y[t] = A[t] \left(Q'x[t] + q' + \lambda' \sum_{i=1}^m \frac{1}{d_i - C_i x[t]} C_i^\top \right), \forall t = 0, \dots, k \right\}.$$

Now, we may prove the following impossibility results.

Theorem 5. *The cardinality of the set $\mathcal{M}[k]$ is uncountably infinite for all $k \in \mathbb{Z}_{\geq 0}$.*

Proof. The proof is similar to that of Theorem 3. □

5. Conclusions

In this paper, we studied the problem of how to keep a utility function confidential even when the network over which the function is being optimized is compromised. Particularly, we considered the problem where an eavesdropper's objective is to reconstruct a quadratic utility function via measuring the decision variable iterations under a gradient method. We considered the impact of different choices of the step size on the reconstructability of the utility function and showed that for the case that the step size is not constant and is selected randomly from a sufficiently large set of appropriate candidates, it is virtually impossible for an eavesdropper to reconstruct the utility function. Therefore, the best design recommendation is to add time-varying agent-dependent random step sizes to the implemented dynamics. In addition to time-varying random step sizes, there are other ingredients that matter, such as having a uniform random direction for the initial condition and not executing too many gradient descent steps (since if the number of steps is below a threshold the solution cannot be uniquely determined even if having access to extraordinary computational capabilities). An interesting avenue for future research could be to devise a tractable algorithm that can approximate the utility function and bound the accuracy of the approximation based on the statistics of the step size selection method. This can be done by using set-membership identification techniques (see Remark 7) for bounding the difference between the identified and the true set of permissible parameters. Another avenue for future research could be to also study quasi-Newton or Newton methods because they require fewer iterations for converge and, thus, potentially minimize the amount of the leaked information.

References

- [1] B. A. Akyol, Cyber security challenges in using cloud computing in the electric utility industry, Tech. Rep. PNNL-21724, Pacific Northwest National Laboratory, Richland, Washington 99352 (September 2012).
- [2] M. Bishop, Computer security: Art and science, Addison-Wesley, 2002.
- [3] Y. Chen, V. Paxson, R. H. Katz, Whats new about cloud computing security, Tech. Rep. UCB/EECS-2010-5, University of California, Berkeley (January 2010).
- [4] V. Cevher, S. Becker, M. Schmidt, Convex optimization for big data: Scalable, randomized, and parallel algorithms for big data analytics, Signal Processing Magazine, IEEE 31 (5) (2014) 32–43.
- [5] J. Nocedal, S. J. Wright, Numerical optimization, Springer New York, 1999.

- [6] C. Dwork, Differential privacy: A survey of results, in: *Theory and Applications of Models of Computation*, Springer, 2008, pp. 1–19.
- [7] A. Gupta, K. Ligett, F. McSherry, A. Roth, K. Talwar, Differentially private combinatorial optimization, in: *Proceedings of the 21st Annual ACM-SIAM Symposium on Discrete Algorithms*, 2010, pp. 1106–1125.
- [8] K. Chaudhuri, C. Monteleoni, A. D. Sarwate, Differentially private empirical risk minimization, *Journal of Machine Learning Research* 12 (2011) 1069–1109.
- [9] O. L. Mangasarian, Privacy-preserving linear programming, *Optimization Letters* 5 (1) (2011) 165–172.
- [10] P. Weeraddana, G. Athanasiou, C. Fischione, J. Baras, Per-se privacy preserving solution methods based on optimization, in: *Proceeding of the 52nd IEEE Conference on Decision and Control*, 2013, pp. 206–2011.
- [11] J. Vaidya, H. Yu, X. Jiang, Privacy-preserving SVM classification, *Knowledge and Information Systems* 14 (2) (2008) 161–178.
- [12] J. C. Duchi, M. I. Jordan, M. J. Wainwright, Privacy aware learning, in: F. Pereira, C. J. C. Burges, L. Bottou, K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems* 25, 2012, pp. 1430–1438.
- [13] J. B. Moore, Persistence of excitation in extended least squares, *IEEE Transactions on Automatic Control* 28 (1) (1983) 60–68.
- [14] S. Gentry, V. Saligrama, E. Feron, Dynamic inverse optimization, in: *Proceedings of the American Control Conference*, Vol. 6, 2001, pp. 4722–4727.
- [15] D. P. Bertsekas, J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Athena Scientific Optimization and Computation Series, Athena Scientific, 1997.
- [16] A. A. Goldstein, Convex programming in Hilbert space, *Bulletin of the American Mathematical Society* 70 (5) (1964) 709–710.
- [17] H. Lütkepohl, *Handbook of Matrices*, John Wiley & Sons, 1996.

- [18] R. L. Kosut, M. K. Lau, S. P. Boyd, Set-membership identification of systems with parametric and nonparametric uncertainty, *IEEE Transactions on Automatic Control* 37 (7) (1992) 929–941.
- [19] M. Milanese, A. Vicino, Optimal estimation theory for dynamic systems with set membership uncertainty: An overview, in: M. Milanese, J. Norton, H. Piet-Lahanier, E. Walter (Eds.), *Bounding Approaches to System Identification*, Springer US, 1996, pp. 5–27.
- [20] M. Milanese, C. Novara, Set membership identification of nonlinear systems, *Automatica* 40 (6) (2004) 957–975.
- [21] W. S. Levine, *Control System Fundamentals*, Taylor & Francis, 1999.
- [22] W. S. Levine, *The Control Handbook, Electrical Engineering Handbook*, Taylor & Francis, 1996.
- [23] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

Appendix A. Convergence of Gradient Algorithm with Agent-Dependent Step Sizes

To present the results of this appendix, we need to introduce some notation. For a matrix A , we write $A \leq 0$ if A is a symmetric negative semi-definite matrix. Further, for matrices A and B of the same dimension, we write $A \leq B$ if $A - B \leq 0$.

Note that we follow the update rule $x[k+1] = x[k] - A[k](Qx[k] + q)$, where $A[k]$ is the matrix of step sizes satisfying $c_1 I \leq A[k] \leq c_2 I$ for all k with $0 < c_1 < c_2$. In the reminder of this appendix, we determine conditions on c_1, c_2 that guarantee the convergence of the gradient algorithm. These iterates converge to the maximizer of the utility function so long as they satisfy Wolfe’s conditions; see Theorem 3.2 in [5, p. 38]. From Wolfe’s conditions, for $0 < \epsilon_1 < \epsilon_2 < 1$, we have

$$\begin{aligned} (x[k] - A[k](Qx[k] + q))^T Q(x[k] - A[k](Qx[k] + q)) + q^T [x[k] - A[k](Qx[k] + q)] - x[k]^T Qx[k] - q^T x[k] \\ \leq -\epsilon_1 (Qx[k] + q)^T A[k] (Qx[k] + q), \end{aligned} \quad (\text{A.1a})$$

$$(Qx[k] + q)^T A[k] (Qx[k] - A[k](Qx[k] + q)) + q \leq \epsilon_2 (Qx[k] + q)^T A[k] (Qx[k] + q). \quad (\text{A.1b})$$

We may rewrite (A.1a) and (A.1b), perspectivevely, as

$$\begin{bmatrix} A[k]^{\frac{1}{2}} Qx[k] \\ A[k]^{\frac{1}{2}} q \end{bmatrix}^T \begin{bmatrix} A[k]^{\frac{1}{2}} Q A[k]^{\frac{1}{2}} + (\epsilon_1 - 2)I & \frac{2\epsilon_1 - 3}{2}I + A[k]^{\frac{1}{2}} Q A[k]^{\frac{1}{2}} \\ \frac{2\epsilon_1 - 3}{2}I + A[k]^{\frac{1}{2}} Q A[k]^{\frac{1}{2}} & A[k]^{\frac{1}{2}} Q A[k]^{\frac{1}{2}} + (\epsilon_1 - 1)I \end{bmatrix} \begin{bmatrix} A[k]^{\frac{1}{2}} Qx[k] \\ A[k]^{\frac{1}{2}} q \end{bmatrix} \leq 0,$$

and

$$\begin{bmatrix} A[k]^{\frac{1}{2}} Q x[k] \\ A[k]^{\frac{1}{2}} q \end{bmatrix}^{\top} \begin{bmatrix} I \\ I \end{bmatrix} \left((1 - \epsilon_2)I - A[k]^{\frac{1}{2}} Q A[k]^{\frac{1}{2}} \right) \begin{bmatrix} I \\ I \end{bmatrix}^{\top} \begin{bmatrix} A[k]^{\frac{1}{2}} Q x[k] \\ A[k]^{\frac{1}{2}} q \end{bmatrix} \leq 0.$$

These conditions are satisfied if the following inequalities hold

$$\begin{bmatrix} (\epsilon_1 - 2)I & \frac{2\epsilon_1 - 3}{2}I \\ \frac{2\epsilon_1 - 3}{2}I & (\epsilon_1 - 1)I \end{bmatrix} + \begin{bmatrix} I \\ I \end{bmatrix} (A[k]^{\frac{1}{2}} Q A[k]^{\frac{1}{2}}) \begin{bmatrix} I \\ I \end{bmatrix}^{\top} \leq 0, \quad (\text{A.2a})$$

$$(1 - \epsilon_2)I - A[k]^{\frac{1}{2}} Q A[k]^{\frac{1}{2}} \leq 0. \quad (\text{A.2b})$$

For (A.2a) to hold, it is sufficient to satisfy

$$\begin{bmatrix} (\epsilon_1 - 2)I & \frac{2\epsilon_1 - 3}{2}I \\ \frac{2\epsilon_1 - 3}{2}I & (\epsilon_1 - 1)I \end{bmatrix} + c_2 \lambda_{\max}(Q) \begin{bmatrix} I & I \\ I & I \end{bmatrix} \leq 0, \quad (\text{A.3})$$

where $\lambda_{\max}(Q)$ denotes the largest eigenvalue of Q . Using Schur's complement, we can translate (A.3) to

$$\begin{aligned} \epsilon_1 - 2 + c_2 \lambda_{\max}(Q) &< 0, \\ \left(\epsilon_1 - 1 + c_2 \lambda_{\max}(Q) \right) \left(\epsilon_1 - 2 + c_2 \lambda_{\max}(Q) \right) - \left(\epsilon_1 - \frac{3}{2} + c_2 \lambda_{\max}(Q) \right)^2 &< 0. \end{aligned}$$

This holds if $\epsilon_1 - 2 + c_2 \lambda_{\max}(Q) < 0$. For (A.2b) to hold, it is sufficient to have $1 - \epsilon_2 - c_1 \lambda_{\min}(Q) \leq 0$, where $\lambda_{\min}(Q)$ denotes the smallest eigenvalue of Q . Therefore, for the gradient algorithm to converge, it suffices to select $c_1 > (1 - \epsilon_2)/\lambda_{\min}(Q)$ and $c_2 < (2 - \epsilon_2)/\lambda_{\max}(Q)$ for $0 < \epsilon_1 < \epsilon_2 < 1$.